



## Genomics in rugby union: a review and future prospects

Journal:	<i>European Journal of Sports Science</i>
Manuscript ID:	TEJS-2014-0699.R1
Manuscript Type:	Review
Keywords:	Genetics, Musculoskeletal, Physiology, Team Sport

SCHOLARONE™  
Manuscripts

**Genomics in rugby union: a review and future prospects**

**Abstract**

This article introduces some aspects of sports genomics in a rugby union context, considers the rugby-specific genetic data in the published literature and outlines the next research steps required if the potential applications of genetic technology in rugby union, also identified here, are to become possible. A substantial proportion of the inter-individual variation for many traits related to rugby performance, including strength, short-term muscle power,  $VO_{2max}$ , injury susceptibility and the likelihood of being an elite athlete is inherited and can be investigated using molecular genetic techniques. In sports genomics, significant efforts have been made in recent years to develop large DNA biobanks of elite athletes for detailed exploration of the heritable bases of those traits. However, little effort has been devoted to the study of rugby athletes, and most of the little research that has focused on rugby was conducted with small cohorts of non-elite players. With steadily growing knowledge of the molecular mechanisms underpinning complex performance traits and the aetiology of injury, investigating sports genomics in the context of rugby is now a viable proposition and a worthwhile endeavour. The RugbyGene project we describe briefly in this article is a multi-institutional research collaboration in rugby union that will perform molecular genetic analyses of varying complexity. Genetic tests could become useful tools for rugby practitioners in the future and provide complementary and additional information to that provided by the non-genetic tests currently used.

**Key words;** sports genetics, athlete status, rugby physiology

**Introduction**

The majority of scientific investigation into player performance in rugby union has focused on environmental factors such as training methods, dietary supplementation and recovery strategies (Barr, Sheppard, Gabbett, & Newton, 2014; Bradley et al., 2014; West et al., 2013), with a great deal of epidemiological research also investigating injury frequency and risk including a recent focus on brain injury (Gardner, Iverson, Williams, Baker, & Stanwell, 2014; Raftery, 2014). However,

considerable evidence shows that performance and injury traits are highly (but variably) heritable (Bouchard et al., 1999; De Moor et al., 2007; Hakim, Cherkas, Spector, & MacGregor, 2003; Peeters et al., 2007; Simoneau & Bouchard, 1995), yet little scientific effort has been invested to elucidate this inter-individual variation within rugby union. Rugby union athletes are qualitatively and quantitatively different from other athlete groups (both individual and team), in that there are vast difference in the physiological and anthropometric characteristics across a single rugby union team according to playing position (Smart, Hopkins, & Gill, 2013). Rugby union is also distinctive in that individual clusters of positions require different movement patterns in elite players (Quarrie, Hopkins, Anthony, & Gill, 2012) and thus differ in their metabolic demands.

For example, anthropometric and physiological variables differ significantly across playing positions, with elite scrum half players averaging ~177 cm and ~85 kg, in contrast to props averaging ~185 cm and ~117 kg – a difference of 8 cm and 32 kg (Fuller, Taylor, Brooks, & Kemp, 2013). Furthermore, in terms of positional specific physiological differences that may be reflected in players' genetic variation, backs show lower maximal strength compared to forwards in terms of bench press (difference ~11 kg), back squat (difference ~18 kg) and power clean (difference ~9 kg; Smart et al., 2011). However, backs are faster sprinting 10 m (difference = ~0.09 s) and 20 m (difference = ~0.11 s) than forwards (Smart, Hopkins, Quarrie, & Gill, 2011) and these differences become larger when specific positions are considered (Smart et al., 2013). Positional differences are further evidenced by game demand data that shows the requirement for differing metabolic capacities dependent on playing position. Backs travel ~12% greater total distance (backs 6545 m, forwards 5850 m), achieving maximum speeds ~15% greater (forwards 26.3 km.h<sup>-1</sup>, backs 30.4 km.h<sup>-1</sup>) and are engaged in two to three times more high-intensity running than forwards (Cahill, Lamb, Worsfold, Headey, & Murray, 2012; Roberts, Trewartha, Higgitt, El-Abd, & Stokes, 2008). Furthermore, forwards perform more (~10%) high-intensity static exertion activities (rucks, mauls, scrums and line-outs) than backs (1.6%) and spend ~20% more time running faster than 80% of their maximal speed (Cahill et al., 2012; Roberts et al., 2008) (Cahill et al., 2012; Roberts et al., 2008, respectively). The back three players (wings and fullbacks) travel the greatest distance at the lowest movement speeds (0-2 m.s<sup>-1</sup>) but also

at the highest speeds ( $>8 \text{ m.s}^{-1}$ ). In contrast, front row forwards (props and hookers) travel the greatest distance at moderately low (but not the lowest) movement speeds ( $2\text{-}4 \text{ m.s}^{-1}$ ) and the least distance at high speeds ( $>8 \text{ m.s}^{-1}$ : Quarrie et al., 2012). These movement patterns correspond with the metabolic demands necessary to perform at the highest level in a given playing position within rugby union. The back three movement pattern suggests a requirement for a high proportion of fast twitch muscle fibres and highly developed anaerobic energy metabolism, with the front row movement pattern suggesting a greater capacity for rapid recovery between high-intensity static exertion activities. Given the highly heritable nature of these and other physical phenotypes that contribute to success in a given playing position, the differences should be reflected in distinct genetic characteristics. Furthermore, since rugby union athletes perform under a well-defined set of rules and parameters, which are ubiquitous across all playing positions, they present an ideal cohort via which to study the importance of genetic variation in sport.

In the present review, we will introduce the field of sports genomics, briefly review the current genetic research available for consideration within rugby union and outline the steps necessary to progress genomic understanding and applications in rugby union. Specifically, we will discuss some candidate genes that have been associated with physiological and anthropometric characteristics, injury risk, skill acquisition and athlete status. We will also introduce the highly collaborative RugbyGene project.

**Human Genomic variation**

The variation that exists in the human genome has only relatively recently be documented via progress of the Human Genome Project. Initially, ~11 million single nucleotide polymorphisms (SNP's) and 3 million short insertions and deletions were identified (Frazer et al., 2007; Sachidanandam et al., 2001; The International HapMap Consortium, 2005). and, at the time of writing, more than 88 million SNP's have been validated ([http://www.ncbi.nlm.nih.gov/projects/SNP/snp\\_summary.cgi](http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi)).The vast majority of the ~3 billion nucleotides that constitute a human genome do not code for proteins, although most of those nucleotides nonetheless contribute to biological function in some manner via regulation of gene

expression (ENCODE Project Consortium, 2012). There are a number of different types of human genomic variation that affect biological function, including rare mutations, structural variations and common SNP's.

The loss of function nucleotide change in the myostatin (*MSTN*) gene is an example of a rare mutation of relevance to human physiological function and results in an absence of the myostatin protein, which is a negative regulator of muscle growth. In humans, only one report of this mutation exists, causing extreme hyper-muscularity and a lean phenotype. At six days old, the affected child showed twice the quadriceps muscularity (cross-sectional area) than that of an age-matched control (6.7 cm<sup>2</sup> compared to 3.1 cm<sup>2</sup>; Schuelke et al., 2004). The negative impacts (if any) of this mutation have not yet been reported and at 4.5 years the child showed no pathological symptoms. However, because myostatin affects other muscle types including cardiac muscle, later-onset pathological symptoms resulting from excessive myocardial growth are a possibility.

Genetic variations where an allele occurs relatively commonly ( $\geq 1\%$ ) are called polymorphisms, as opposed to mutations. An example of one such polymorphism relevant to exercise physiology is the insertion/deletion (I/D) variation on intron 16 of the *ACE* gene on chromosome 17. The *ACE* gene encodes the angiotensin converting enzyme, which is the main active product of the rennin-angiotensin system. The insertion of the 287 base pair (bp) DNA sequence (I allele), despite being in a non-protein coding region of the *ACE* gene, produces lower circulating and tissue concentrations of the ACE protein compared to the D allele (Almeida et al., 2010). One main action of the ACE enzyme is to degrade inactive angiotensin I and generate the vasoconstrictor angiotensin II, while another is to degrade vasodilator kinins and thus is hypothesised to influence mitochondrial oxygen consumption and exercise economy in some circumstances (Puthucherry et al., 2011).

An example of a SNP relevant to exercise performance found in an exon (a region of DNA that encodes for protein) is the *ACTN3* R577X gene variant. The *ACTN3* gene encodes for the  $\alpha$ -actinin-3 protein, which is expressed almost exclusively in fast glycolytic type II fibres and is a structural component that binds the actin thin filament to the Z line. This SNP is located on exon 15 of the

*ACTN3* gene (North et al., 1999) and is characterised by the replacement of the normal codon (a three bp sequence that transcribes for an amino acid (Arg; R)) by a premature termination codon (X) at the 577 amino acid position and results in the complete absence of the  $\alpha$ -actinin-3 protein. The absence of this protein (XX genotype) is associated with a lower proportion of type 2 muscle fibres and, accordingly, is found at a lower frequency in elite power/sprint athletes compared to other athletes and non-athletes (Eynon et al., 2013).

**Physiological and anthropometric characteristics**

Classical genetics is the process of estimating the heritability of a given trait and is investigated by studying families, identical twins (monozygotic; MZ) and fraternal twins (dizygotic; DZ). A number of physiological variables have been investigated in this way and are usually the preceding step to investigating specific target genes for further analysis. For example, Simoneau and Bouchard (1995) showed the genetic heritability of muscle fibre type proportion was ~50% when investigating 58 DZ and 35 MZ twin pairs. Furthermore, following 20 weeks of endurance training in 98 two-generation families (total n = 481), the heritable component of the ability to adapt, specifically an increase  $VO_{2max}$ , was ~50% in the well-known HERITAGE family study (Bouchard et al., 1999). While these heritability estimates are substantial, other anthropometric phenotypes are estimated to be even greater: for example BMI at ~60%, height at ~80% (Silventoinen, Magnusson, Tynelius, Kaprio, & Rasmussen, 2008) and mesomorphy ~80% (Peeters et al., 2007).

Subsequently, molecular genetic technologies have been applied to examine the specific genes and their variations that might explain the heritable component of these and other phenotypes. For example, two intronic SNPs in the *TRHR* gene (rs7832552 and rs16892496) collectively account for ~5 kg of the variability in lean body mass (Liu et al., 2009). The *TRHR* gene encodes for the thyrotrophin-releasing hormone receptor that when bound to thyrotrophin-releasing hormone, results in a signalling cascade which leads to the release of thyroxine, an important component in the development of mammalian skeletal muscle. This gene variant, especially in combination with many others, could help explain a notable proportion of the ~80% heritability of mesomorphy mentioned previously and could be applied to the assessment of rugby players. Similarly, with regards to  $VO_{2max}$ ,

Bouchard et al. (2011) found the *PRDM1* gene accounted for ~7% of the inter-individual variability in adaptation to endurance training. The protein encoded by *PRDM1* is widely expressed and has been implicated in skeletal muscle fibre type differentiation (Beermann, Ardelt, Girgenrath, & Miller, 2010). The biology explaining why this gene variant is associated with  $VO_{2max}$  has yet to be elucidated but, again in combination with other variants such as those identified by Bouchard et al. (2011), could provide a useful tool in rugby.

Additionally, the *ACTN3 R577X* gene variant (already mentioned) could, when combined with other genetic variants, be a molecular marker for muscle fibre type proportion. Vincent et al. (2007) showed that the RR genotype (presence of the protein) had ~5% greater type IIx fibres than XX genotype individuals (absence of the protein) and had ~5% larger relative surface area per fibre type ( $n = 43$ ). Supporting evidence followed using a larger sample ( $n = 94$ ) where the XX genotype was associated with ~3% greater type I muscle fibre type proportion than RR individuals (Ahmetov et al., 2011).

### **Injury risk**

Possibly the most useful application of predictive genetics within sports genomics will be in the field of injury risk and severity estimation. Hakim et al. (2003) examined frozen shoulder (FS) and tennis elbow (TE) in 865 MZ and 963 DZ twin pairs and reported 42% heritability for FS and 40% for TE. Additionally, while we are not aware of heritability estimates for mild-to-moderate traumatic brain injury (TBI) such as concussion, the likelihood of suffering more severe future neurological disorders following these injuries is significant (Goldman et al., 2006).

Probably the most explored gene regarding tendon and ligament injuries is *COL5A1*, which encodes for the  $\alpha 1$  chain of type V collagen, a minor fibrillar collagen (Hildebrand, Frank, & Hart, 2004). Mokone et al. (2006) investigated the difference in genotype frequency in South African Caucasians between 111 patients with Achilles tendon pathology and 129 control subjects. They showed that the CC genotype was over-represented in the control population, suggesting a protective role of the C allele against tendon injury. A similar association, indicated a protective C allele, has been shown with cruciate ligament injuries in South African women (Posthumus et al., 2009). Assessment of this

*COL5A1* genetic marker, in combination with other markers yet to be identified, might provide a useful tool in rugby for individualising training load and mode to reduce incidence of injury.

The *APOE* gene is on chromosome 19 and encodes apolipoprotein E-based peptide (ApoE), a candidate marker for risk and severity of traumatic brain injury. ApoE is a protein that plays a pivotal role in cholesterol metabolism and has been linked to neurobiological function and, specifically, susceptibility to late-onset and sporadic Alzheimer’s disease via *APOE* gene polymorphisms (Bu, 2009). The *APOE* E4 allele has also been associated with neurodegenerative cascade subsequent to TBI and the severity of axonal injury in mouse models (Bennett et al., 2013), with human studies showing an association between the *APOE* E4 allele and poor outcome following TBI (Laskowitz & Vitek, 2007). Functionally, carriers of the *APOE* E4 allele have presented with reduced motor rehabilitation outcomes, poorer neurocognitive outcomes, increased cognitive impairments and memory defects following TBI (Reviewed in, Gokhale & Laskowitz, 2013). This genetic marker, combined with others such as *TNFA* -308, a promoter region within *GRIN2A*, and others including more yet to be identified, could improve estimation of both recovery duration from TBI/concussion in rugby and the risk of longer-term neurocognitive problems – potentially useful from both short-term player management and long-term player health perspectives. The most interesting, but preliminary, investigation to date observed the recovery-time of 51 athletes following medically diagnosed concussion: McDevitt et al. (2014) investigated a repeat polymorphism in the promoter region within the *GRIN2A* gene and found the chance of prolonged recovery from concussion was 4.3 times greater for homozygous risk allele carriers.

**Skill acquisition and cognitive ability**

In a large study of 11,000 twin pairs from four countries (US, AUS, UK and The Netherlands), the heritability of general cognitive ability was found to increase linearly from childhood (41%) to adulthood (66%; Haworth et al., 2010), while the ability to learn motor skills is also highly heritable at ~70% (Fox, Hershberger, & Bouchard, 1996). One candidate to explain some of this heritability is the *BDNF* gene, which encodes for brain-derived neurotrophic factor, a protein that influences cortical synaptic plasticity (Akaneya, Tsumoto, Kinoshita, & Hatanaka, 1997). Individuals possessing the Met



allele of the Val66Met polymorphism in the *BDNF* gene show lower increases in motor evoked potentials after motor training (Kleim et al., 2006). Furthermore, carriers of the Met allele showed no change in neurological excitability from transcranial magnetic stimulation (TMS), whereas carriers of the Val allele showed a 67% increase (Missitzi et al., 2011). These data suggest that carriers of the Met allele might show lower adaptation to motor learning through a lack of neurobiological excitability, possibly related to altered cortical synaptic plasticity. Another candidate is the Val158Met polymorphism in the catechol-O-methyl transferase (*COMT*) gene that encodes the COMT enzyme which is a major catabolising enzyme of the dopamine pathway. Individuals homozygous for the Met allele show impaired inhibition of prepotent responses (i.e. the ability to suppress automated behavioural responses) and reduced working memory plasticity (Bellander et al., 2014). It would be fascinating to test the hypothesis that favourable alleles of these *BDNF* and *COMT* gene variations, and others related to neurobiological function, may be found more frequently in playing positions with greater requirement to execute relatively fine motor skills and greater responsibility for making tactical decisions in rugby such as the half backs. Genetic tests for skill acquisition and cognitive ability might then become useful additions to other non-genetic assessments in talent identification and individualisation of skill and tactical elements of training programmes.

#### **Athlete status**

De Moor et al. (2007) investigated the genetic component of athlete status in female 1,000 DZ and 793 MZ twins and reported a heritability estimate of ~70% for athlete status. We are not aware of similar data for males. In a review published in 2012, at least 79 genetic markers were identified that had been associated, in at least one prior research paper, with elite athlete status (Ahmetov & Fedotovskaya, 2012). However, that number was reduced to 20 when the criterion was at least two prior research papers, and probably even some of those associations will not prove to be true as more data are accumulated. The first scientific investigation to assess the molecular genetic component of elite athlete status (Gayagay et al., 1998) showed a significant association of the *ACE* I/D variant (mentioned previously) with elite status in 64 rowers. Subsequently, the *ACE* I allele has been associated with elite performance in a variety of sports, though not consistently, and the research is probably best summarised by reference to the meta-analysis of Ma et al. (2013) who found that the II

genotype of *ACE* I/D was associated with physical performance (odds ratio 1.23), especially endurance performance (odds ratio 1.35).

For the *ACTN3* R577X variant already mentioned, the R allele has been consistently associated with elite power and sprint athletes from a variety of backgrounds and in top elite sprinters, a complete absence of the XX genotype has been identified (Niemi & Majamaa, 2005; Yang et al., 2003), compared to an XX genotype frequency of ~18% in the general Caucasian population. Again, the meta-analysis by Ma et al. (2013) nicely summarises the association of the R allele with elite power athlete status (odds ratio 1.21).

Gradually, a number of athlete cohorts have emerged, hosting steadily larger samples of elite and sub-elite athletes from various sports for the investigation of athlete status (Pitsiladis et al., 2013). We are also aware that a UK athlete cohort is being established. One potential application of knowledge about the genetic characteristics of elite athletes, including in rugby union, could be an enhancement to talent identification programmes that are in some cases already quite sophisticated.

**Genetics and rugby union**

As early as 1922 scientists were hypothesising about the role of heritability in rugby union athletes. Jack (1922) documented the playing positions in 23 sets of elite rugby-playing brothers (n = 63), including a number of international representatives, and concluded that “*the ability required for playing in certain positions in rugby football is inherited*”. We are aware of only three full publications in peer-reviewed journals applying molecular genetics to rugby union. Goh et al. (2009) reported that the II *ACE* genotype was associated with a higher ventilatory threshold in non-elite Asian rugby players but the very small cohort (n = 17) is a major limitation to that study. Bell et al. recently assessed *ACE* I/D and *ACTN3* R577X genotypes in 68 and 102 young non-elite rugby union players, respectively (Bell, Colley, Evans, Darlington, & Cooper, 2012; Bell et al., 2010). No associations were identified between either polymorphism and athlete status, playing position or the physiological and anthropometric parameters assessed, perhaps due to the rather small sample size and the sub-elite status of the players. With such little data regarding the genetic characteristics of

rugby union players in the literature, this gap should be filled with high quality data using appropriate experimental designs.

### **Next steps for rugby genomics research**

#### **Cohort**

For sports genomics research to be truly relevant to the preparation and management of elite competitors, the athletes from which the initial data are derived must themselves be considered elite. In the context of rugby union, we propose a definition of ‘elite’ as athletes competing in the highest competitive league of a ‘Tier 1’ rugby nation. Comparisons between sub-elite (e.g. the next tier in the league structure), suitably matched controls and elite athletes would also be useful for explanatory purposes. Given the evolving nature of elite rugby, the era in which athletes competed at an elite level also needs to be defined. Rugby union has changed dramatically in the >100 years of its existence and that change has certainly continued even since the sport turned professional ~20 years ago. Nevertheless, we propose that 1995 onwards is a playing era inclusion criterion that can sensibly be justified.

Another critical characteristic of the cohort is that it is large in number. Genetic research requires a large sample size to obtain sufficient statistical power and simultaneously minimise the likelihood of type 1 statistical error (false positive associations). Only by using a large cohort can patterns and associations in the data be identified reliably and sound conclusions made regarding associations of genetic variations within a given cohort. Consequently, large numbers (many hundreds, and preferably thousands) of ‘Tier 1’ rugby athletes are required – and that will only be achieved through international research collaboration. Additionally, geographic ancestry is an important consideration for case-control and genotype-phenotype association study designs and therefore analysis of molecular genetic markers should preferably be performed on athletes from a well-defined geographic ancestry cluster. A more difficult aim to achieve would be to recruit large numbers of players from all geographic ancestry clusters commonly found in rugby union, although this would be a very powerful approach scientifically.

**Genotype-phenotype associations that should be investigated**

Following the recruitment of the appropriate cohort, the next step would be to investigate genotype associations with a range of phenotypes such as success in rugby (elite athlete status and playing position), physiological, anthropometric and other performance variables, incidence of injury, etc. We recently presented preliminary results involving 272 elite (as defined above) rugby union athletes (compared to 141 controls matched for geographic ancestry) at a scientific conference, focussing on elite athlete status in that report. Specifically, we reported an association between the *ACTN3* R577X gene variant and elite rugby union athlete status as well as playing position (Heffernan et al., 2014). We are continuing to build this cohort in what we have termed the RugbyGene project.

Identifying genetic associations (in a cohort of elite rugby athletes) with rugby-specific physiological and anthropometric variables for example those aspects of strength, speed and body composition assessed by Smart et al. (2013), would further exemplify the importance of the genetic component to rugby. Furthermore, rugby union has one of the highest reported incidents of match play injuries in all professional team sports (Brooks & Kemp, 2008), with an injury incidence of 81 per 1000 match play hours resulting in ~20 days absence per injury (Williams, Trewartha, Kemp, & Stokes, 2013). Research collaborations that combine these kinds of large, meticulously-collected rugby injury databases with genetic analyses conducted on those very same players, could be extremely fruitful in explaining some of the as yet unexplained inter-individual variability in injury susceptibility and could identify novel markers of injury risk in rugby.

Concussion risk in rugby, as well as consideration of the potential longer-term consequences, is clearly and justifiably a topic of much attention at this time (Raftery, 2014). Accordingly, Gardner et al. (2014) recently conducted a comprehensive meta-analysis showing that in men's rugby union, the incidence rate of concussion was ~4.7 per 1000 match play hours. Data from 16 studies showed that at the elite/international level the incidence rate was lower (~1.2 per 1000 match play hours), though still considerable (Gardner et al., 2014). These notable injury rates make research efforts to identify molecular markers for the risk of specific rugby-related injuries, such as concussion, highly warranted. Indeed, the development of tools to identify individuals at greater risk of concussion in

rugby and greater risk of longer-term pathological neurobiological changes following a career playing rugby would seem highly responsible in the context of player welfare.

### **Advanced genomics technology**

For rugby union as in other areas within sports genomics, the starting point for the investigation of genetic variation is through the candidate gene (hypothesis driven) approach including the genes mentioned here in this manuscript and many others. This approach is the process of considering the biological mechanisms of a given trait (for example, incidence of tendinopathy or high  $VO_{2max}$ ) and investigating previously identified genetic variants within genes known or suspected to affect the relevant biological pathways in some way. While the candidate gene approach is by far the most utilised technique in quantifying molecular genetic markers of sport-related phenotypes and a good starting point, other more complex analysis techniques are required for this field to realise its full potential (Pitsiladis et al., 2013).

The hypothesis-free approach of conducting genome wide association studies (GWAS) has been utilised extensively to identify new genetic variants in various domains within human biology (Wolfarth et al., 2014) and is recommended for identifying novel genetic variants in rugby union. GWAS is the process of investigating large numbers of known SNP's simultaneously (~2 million, for example) for a given complex trait (Visscher, Brown, McCarthy, & Yang, 2012). As already mentioned, complex traits of relevance to rugby union could be sprinting ability, muscle strength, incidence of tendinopathy or simply being an elite rugby union player (elite athlete status). Importantly, given the large number of hypotheses tested statistically, only the strongest associations are usually accepted to be true results (e.g. when  $P < 5 \times 10^{-8}$ ), although a very large cohort size and/or strong genetic effect sizes are usually required for this approach to be effective. A strength of any hypothesis-free approach like GWAS is that new variants which reveal new biological insight can be discovered - and then further investigated experimentally.

Genetic testing technologies have advanced to such an extent that investigating all nucleotides in a gene, all protein-coding genes, or even the whole genome is now possible via direct sequencing. Depending on the availability of funding, participants and other resources including laboratory

equipment and bioinformatics expertise, multiple options are available to exploit these techniques for the identification of novel mutations, polymorphisms or structural variants. Firstly, targeting specific genes associated with a given phenotype, such as the *MSTN* gene and muscle mass, and sequencing every nucleotide (~7,000 bp) of that gene in a large cohort of elite rugby union players for whom strength and muscle mass are also known could be an elegant approach to identifying novel genetic variants associated with muscle size and strength in rugby players. To examine larger regions than one gene, one could utilise whole-exome sequencing, which uses the same rational as targeted gene sequencing but to a much greater extent, targeting all protein-coding DNA sequences (~230,000 exons or ~30 million bp; ~1% of the human genome). Finally, providing that considerable resources including an appropriately large cohort were available, assessment of the whole-genome (~3 billion bp) would be the ultimate, most comprehensive method of identifying novel mutations or polymorphisms of functional importance in rugby union athletes. Currently, whole-genome-sequencing is not recommended for reasons of cost, logistics, interpretation and statistics, although the eventual application to rugby is inevitable.

Following the identification of novel gene variations through the methods described above, simple mathematical models can be applied to simultaneously quantify the effect of several gene variants. Williams and Folland (2008) developed the total genotype score (TGS) as a way to represent the simultaneous influence of multiple genetic markers as a simple value that is intuitively understandable. According to the TGS, all scores lie between 0-100 and a higher score indicates a greater genetic suitability for a given phenotype such as muscle mass, sprinting speed, reduced injury risk, greater cognitive ability – depending on the genetic variants that are used to calculate the specific TGS in each case. Thus, multiple TGS scores, each directed towards a given phenotype, could be calculated for each individual. Subsequently, both hypothetical (Hughes, Day, Ahmetov, & Williams, 2011) and experimental (Eynon, Ruiz, Meckel, Morán, & Lucia, 2011; Ruiz et al., 2010; Ruiz et al., 2009; Santiago et al., 2010) studies have used this method in attempts to elucidate the ‘optimal polygenetic profile’ for a given cohort with some limited success. Other statistical methods to capture the simultaneous influence of multiple genetic markers include construction of inferred pseudo-

haplotypes when appropriate or simply calculating the number of favourable alleles carried by an individual across a number of unrelated genetic loci.

### Conclusion

To generate useful knowledge about the genomics of rugby union, hundreds or thousands of individual players are required in a research cohort for one to be confident that the data are robust - particularly when performing some of the more complex genomic analyses outlined in the preceding section. We and others are collaborating in a large, multi-institutional effort to make progress in this field - the RugbyGene project. There will need to be a simultaneous analysis of genomic and phenotypic (performance and injury) data to maximise the potential of this kind of research. The challenge will then be to apply genomic technologies, alongside existing non-genomic data of course, to personalise the management of players in elite rugby union and facilitate the prescription of training, nutrition, playing load and management of injury risk in a more individualised manner than is currently possible to improve both performance and player welfare.

### References

- Ahmetov, Ildus I, Druzhevskaya, Anastasiya M, Lyubaeva, Ekaterina V, Popov, Daniil V, Vinogradova, Olga L, & Williams, Alun G. (2011). The dependence of preferred competitive racing distance on muscle fibre type composition and ACTN3 genotype in speed skaters. *Experimental Physiology*, 96(12), 1302-1310.
- Ahmetov, Ildus I, & Fedotovskaya, Olga N. (2012). Sports genomics: Current state of knowledge and future directions. *Cellular and Molecular Exercise Physiology*, 1(1), e1.
- Akaneya, Yukio, Tsumoto, Tadaharu, Kinoshita, Shuichiro, & Hatanaka, Hiroshi. (1997). Brain-derived neurotrophic factor enhances long-term potentiation in rat visual cortex. *The Journal of Neuroscience*, 17(17), 6707-6716.
- Almeida, S. S., Barros, C. C., Moraes, M. R., Russo, F. J., Haro, A. S., Rosa, T. S., . . . Araujo, R. C. (2010). Plasma kallikrein and angiotensin I-converting enzyme N- and C-terminal domain activities are modulated by the insertion/deletion polymorphism. *Neuropeptides*, 44(2), 139-143.
- Barr, M, Sheppard, J, Gabbett, T, & Newton, R. (2014). Long-term training induced changes in sprinting speed and sprint momentum in elite rugby union players. *Journal of Strength and Conditioning Research*, 28(10), 2724-2731.
- Beermann, Mary Lou, Ardel, Magdalena, Girgenrath, Mahasweta, & Miller, Jeffrey Boone. (2010). Prdm1 (Blimp-1) and the expression of fast and slow myosin heavy chain isoforms during avian myogenesis in vitro. *PLoS One*, 5(4), e9951. doi: 10.1371/journal.pone.0009951
- Bell, W., Colley, J. P., Evans, W. D., Darlington, S. E., & Cooper, S. M. (2012). ACTN3 genotypes of rugby union players: distribution, power output and body composition. *Annals of Human Biology*, 39(1), 19-27.
- Bell, W., Colley, J. P., Gwynne, J. R., Glazier, P., Evans, W. D., & Darlington, S. E. (2010). ACE ID genotype and leg power in rugby union players. *Journal of Sports Medicine and Physical Fitness*, 50(3), 350-355.



- Bellander, Martin, Bäckman, Lars, Liu, Tian, Schjeide, Brit-Maren M, Bertram, Lars, Schmiedek, Florian, . . . Lövdén, Martin. (2014). Lower baseline performance but greater plasticity of working memory for carriers of the val allele of the comt val158met polymorphism. *Neuropsychology, Advance online publication*. doi: 10.1037/neu0000088
- Bennett, Rachel E, Esparza, Thomas J, Lewis, Hal A, Kim, Eddie, Mac Donald, Christine L, Sullivan, Patrick M, & Brody, David L. (2013). Human apolipoprotein E4 worsens acute axonal pathology but not amyloid- $\beta$  immunoreactivity after traumatic brain injury in 3xTG-AD mice. *Journal of Neuropathology & Experimental Neurology*, 72(5), 396-403.
- Bouchard, Claude, An, Ping, Rice, Treva, Skinner, James S, Wilmore, Jack H, Gagnon, Jacques, . . . Rao, DC. (1999). Familial aggregation of VO<sub>2</sub> max response to exercise training: results from the HERITAGE Family Study. *Journal of Applied Physiology*, 87(3), 1003-1008.
- Bouchard, Claude, Sarzynski, Mark A, Rice, Treva K, Kraus, William E, Church, Timothy S, Sung, Yun Ju, . . . Rankinen, Tuomo. (2011). Genomic predictors of the maximal O<sub>2</sub> uptake response to standardized exercise training programs. *Journal of Applied Physiology*, 110(5), 1160-1170.
- Bradley, W., Cavanagh, B., Douglas, W., Donovan, T. F., Morton, J. P., & Close, G. L. (2014). Quantification of training load, energy intake and physiological adaptations during a rugby pre season: a case study from an elite European rugby union squad. *Journal of Strength Conditioning Research, Advance online publication*. doi: 10.1519/jsc.0000000000000631
- Brooks, John HM, & Kemp, Simon. (2008). Recent trends in rugby union injuries. *Clinics in Sports Medicine*, 27(1), 51-73.
- Bu, Guojun. (2009). Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. *Nature Reviews Neuroscience*, 10(5), 333-344.
- Cahill, Nicola, Lamb, Kevin, Worsfold, Paul, Headey, Roy, & Murray, Stafford. (2012). The movement characteristics of English premiership rugby union players. *Journal of Sports Sciences*, 31(3), 229-237.
- De Moor, Marleen HM, Spector, Tim D, Cherkas, Lynn F, Falchi, Mario, Hottenga, Jouke Jan, Boomsma, Dorret I, & De Geus, Eco JC. (2007). Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Research and Human Genetics*, 10(06), 812-820.
- ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57-74.
- Eynon, Nir, Hanson, Erik D, Lucia, Alejandro, Houweling, Peter J, Garton, Fleur, North, Kathryn N, & Bishop, David J. (2013). Genes for elite power and sprint performance: ACTN3 leads the way. *Sports Medicine*, 43(9), 803-817.
- Eynon, Nir, Ruiz, Jonatan R, Meckel, Yoav, Morán, María, & Lucia, Alejandro. (2011). Mitochondrial biogenesis related endurance genotype score and sports performance in athletes. *Mitochondrion*, 11(1), 64-69.
- Fox, Paul W, Hershberger, Scott L, & Bouchard, Thomas J. (1996). Genetic and environmental contributions to the acquisition of a motor skill. *Nature*, 28(384), 356-358.
- Frazer, Kelly A, Ballinger, Dennis G, Cox, David R, Hinds, David A, Stuve, Laura L, Gibbs, Richard A, . . . Leal, Suzanne M. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164), 851-861.
- Fuller, C. W., Taylor, A. E., Brooks, J. H., & Kemp, S. P. (2013). Changes in the stature, body mass and age of English professional rugby players: a 10-year review. *Journal of Sports Sciences*, 31(7), 795-802.
- Gardner, Andrew J, Iverson, Grant L, Williams, W Huw, Baker, Stephanie, & Stanwell, Peter. (2014). A systematic review and meta-analysis of concussion in rugby union. *Sports Medicine*, 1-15. doi: 10.1007/s40279-014-0233-3
- Gayagay, George, Yu, Bing, Hambly, Brett, Boston, Tanya, Hahn, Alan, Celermajer, David S, & Trent, Ronald J. (1998). Elite endurance athletes and the ACE I allele—the role of genes in athletic performance. *Human Genetics*, 103(1), 48-50.



- Goh, KP, Chew, K, Koh, A, Guan, M, Wong, YS, & Sum, CF. (2009). The relationship between ACE gene ID polymorphism and aerobic capacity in Asian rugby players. *Singapore Medical Journal*, 50(10), 997-1003.
- Gokhale, Sankalp, & Laskowitz, Daniel T. (2013). ApoE and outcome after traumatic brain injury. *Clinical Lipidology*, 8(5), 561-571.
- Goldman, Samuel M, Tanner, Caroline M, Oakes, David, Bhudhikanok, Grace S, Gupta, Anjali, & Langston, J William. (2006). Head injury and Parkinson's disease risk in twins. *Annals of Neurology*, 60(1), 65-72.
- Hakim, AJ, Cherkas, LF, Spector, TD, & MacGregor, AJ. (2003). Genetic associations between frozen shoulder and tennis elbow: a female twin study. *Rheumatology*, 42(6), 739-742.
- Haworth, CMA, Wright, MJ, Luciano, M, Martin, NG, De Geus, EJC, Van Beijsterveldt, CEM, . . . Davis, OSP. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, 15(11), 1112-1120.
- Heffernan, S. M., Kilduff, L. P., M., Erskine R., Day, S. H., McMahon, G. E., Cook, C. J., . . . Williams, A. G. (2014). ACE and ACTN3 gene variants are associated with elite status and playing position in rugby union. In Bases 2013 abstracts. *Journal of Sports Sciences*, 32(Suppl 1), s49-s50. doi: 10.1080/02640414.2014.896604
- Hildebrand, KA, Frank, CB, & Hart, DA. (2004). Gene intervention in ligament and tendon: current status, challenges, future directions. *Gene Therapy*, 11(4), 368-378.
- Hughes, David C, Day, Stephen H, Ahmetov, Ildus I, & Williams, Alun G. (2011). Genetics of muscle strength and power: polygenic profile similarity limits skeletal muscle performance. *Journal of Sports Sciences*, 29(13), 1425-1434.
- Jack, HW. (1922). Brothers in rugby football: More evidence that brothers are selected to play in the same or similar positions on football teams. *Journal of Heredity*, 13(4), 161-162.
- Kleim, Jeffrey A, Chan, Sheila, Pringle, Erin, Schallert, Kellan, Procaccio, Vincent, Jimenez, Richard, & Cramer, Steven C. (2006). BDNF val66met polymorphism is associated with modified experience-dependent plasticity in human motor cortex. *Nature Neuroscience*, 9(6), 735-737.
- Laskowitz, Daniel T, & Vitek, Michael P. (2007). Apolipoprotein E and neurological disease: therapeutic potential and pharmacogenomic interactions. *Pharmacogenomics*, 8, 959-969.
- Liu, Xiao-Gang, Tan, Li-Jun, Lei, Shu-Feng, Liu, Yong-Jun, Shen, Hui, Wang, Liang, . . . Chen, Xiang-Ding. (2009). Genome-wide association and replication studies identified TRHR as an important gene for lean body mass. *The American Journal of Human Genetics*, 84(3), 418-423.
- Ma, Fang, Yang, Yu, Li, Xiangwei, Zhou, Feng, Gao, Cong, Li, Mufei, & Gao, Lei. (2013). The association of sport performance with ACE and ACTN3 genetic polymorphisms: A systematic review and meta-analysis. *PloS One*, 8(1), e54685.
- McDevitt, Jane, Tierney, Ryan, Phillips, Jacqueline, Gaughan, John, Torg, Joseph S, & Krynetskiy, Evgeny. (2014). Supporting the Concept of Genetic Predisposition to Prolonged Recovery Following a Concussion. *Orthopaedic Journal of Sports Medicine*, 2(2 suppl). doi:10.1177/2325967114S00078
- Missitzi, Julia, Gentner, Reinhard, Geladas, Nickos, Politis, Panagiotis, Karandreas, Nikos, Classen, Joseph, & Klissouras, Vassilis. (2011). Plasticity in human motor cortex is in part genetically determined. *The Journal of Physiology*, 589(2), 297-306.
- Mokone, GG, Schweltnus, MP, Noakes, TD, & Collins, M. (2006). The COL5A1 gene and Achilles tendon pathology. *Scandinavian Journal of Medicine & Science in Sports*, 16(1), 19-26.
- Niemi, Anna-Kaisa, & Majamaa, Kari. (2005). Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *European Journal of Human Genetics*, 13(8), 965-969.
- North, Kathryn N, Yang, Nan, Wattanasirichaigoon, Duangrurdee, Mills, Michelle, Easteal, Simon, & Beggs, Alan H. (1999). A common nonsense mutation results in  $\alpha$ -actinin-3 deficiency in the general population. *Nature Genetics*, 21(4), 353-354.

Peeters, MW, Thomis, MA, Loos, RJF, Derom, CA, Fagard, Robert, Claessens, AL, . . . Beunen, GP. (2007). Heritability of somatotype components: a multivariate analysis. *International Journal of Obesity*(8), 1295-1301.

Pitsiladis, Yannis, Wang, Guan, Wolfarth, Bernd, Scott, Robert, Fuku, Noriyuki, Mikami, Eri, . . . Lucia, Alejandro. (2013). Genomics of elite sporting performance: what little we know and necessary advances. *British Journal of Sports Medicine*, 47(9), 550-555.

Posthumus, Michael, September, Alison V, Keegan, Michael, O’Cuinneagain, Dion, Van der Merwe, Willem, Schwellnus, Martin P, & Collins, Malcolm. (2009). Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant. *British Journal of Sports Medicine*, 43(5), 352-356.

Puthuchery, Zudin, Skipworth, James R A, Rawal, Jai, Loosemore, Mike, Van Someren, Ken, & Montgomery, Hugh E. (2011). The ACE gene and human performance. *Sports Medicine*, 41(6), 433-448.

Quarrie, Kenneth L, Hopkins, Will G, Anthony, Mike J, & Gill, Nicholas D. (2012). Positional demands of international rugby union: Evaluation of player actions and movements. *Journal of Science and Medicine in Sport*, 16(4), 353-359.

Rafferty, Martin. (2014). Concussion and chronic traumatic encephalopathy: International Rugby Board’s response. *British Journal of Sports Medicine*, 48(2), 79-80.

Roberts, Simon P, Trewartha, Grant, Higgitt, Rob J, El-Abd, Joe, & Stokes, Keith A. (2008). The physical demands of elite English rugby union. *Journal of Sports Sciences*, 26(8), 825-833.

Ruiz, Jonatan R, Arteta, David, Buxens, Amaya, Artieda, Marta, Gómez-Gallego, Félix, Santiago, Catalina, . . . Lucia, Alejandro. (2010). Can we identify a power-oriented polygenic profile? *Journal of Applied Physiology*, 108(3), 561-566.

Ruiz, Jonatan R, Gómez-Gallego, Félix, Santiago, Catalina, González-Freire, Marta, Verde, Zoraida, Foster, Carl, & Lucia, Alejandro. (2009). Is there an optimum endurance polygenic profile? *The Journal of Physiology*, 587(7), 1527-1534.

Sachidanandam, Ravi, Weissman, David, Schmidt, Steven C, Kakol, Jerzy M, Stein, Lincoln D, Marth, Gabor, . . . Willey, David L. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822), 928-933.

Santiago, C, Ruiz, JR, Muniesa, CA, González-Freire, M, Gómez-Gallego, F, & Lucia, A. (2010). Does the polygenic profile determine the potential for becoming a world-class athlete? Insights from the sport of rowing. *Scandinavian Journal of Medicine and Science in Sports*, 20(1), e188-e194.

Schuelke, Markus, Wagner, Kathryn R, Stolz, Leslie E, Hübner, Christoph, Riebel, Thomas, Kömen, Wolfgang, . . . Lee, Se-Jin. (2004). Myostatin mutation associated with gross muscle hypertrophy in a child. *New England Journal of Medicine*, 350(26), 2682-2688.

Silventoinen, Karri, Magnusson, Patrik KE, Tynelius, Per, Kaprio, Jaakko, & Rasmussen, Finn. (2008). Heritability of body size and muscle strength in young adulthood: a study of one million Swedish men. *Genetic Epidemiology*, 32(4), 341-349.

Simoneau, Jean-Aime, & Bouchard, Claude. (1995). Genetic determinism of fiber type proportion in human skeletal muscle. *The FASEB Journal*, 9(11), 1091-1095.

Smart, Daniel, Hopkins, Will G., Quarrie, Kenneth L., & Gill, Nicholas. (2011). The relationship between physical fitness and game behaviours in rugby union players. *European Journal of Sport Science*, 1-10. doi: 10.1080/17461391.2011.635812

Smart, Daniel J, Hopkins, Will G, & Gill, Nicholas D. (2013). Differences and changes in the physical characteristics of professional and amateur rugby union players. *The Journal of Strength & Conditioning Research*, 27(11), 3033-3044.

The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, 437(7063), 1299-1320.

- Vincent, B., De Bock, K., Ramaekers, M., Van den Eede, E., Van Leemputte, M., Hespel, P., & Thomis, M. A. (2007). ACTN3 (R577X) genotype is associated with fiber type distribution. *Physiological Genomics*, 32(1), 58-63.
- Visscher, Peter M, Brown, Matthew A, McCarthy, Mark I, & Yang, Jian. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, 90(1), 7-24.
- West, D, Cunningham, D, Bevan, H, Crewther, B, Cook, C, & Kilduff, L. (2013). Influence of active recovery on professional rugby union player's ability to harness postactivation potentiation. *The Journal of Sports Medicine and Physical Fitness*, 53(2), 203-208.
- Williams, A. G., & Folland, J. P. (2008). Similarity of polygenic profiles limits the potential for elite human physical performance. *The journal of physiology*, 586(1), 113-121.
- Williams, Sean, Trewartha, Grant, Kemp, Simon, & Stokes, Keith. (2013). A meta-analysis of injuries in senior men's professional rugby union. *Sports Medicine*, 43(10), 1043-1055.
- Wolfarth, Bernd, Rankinen, Tuomo, Hagberg, James M, Loos, RJ, Pérusse, Louis, Roth, Stephen M, . . . Bouchard, Claude. (2014). Advances in exercise, fitness, and performance genomics in 2013. *Medicine and science in sports and exercise*, 46(5), 851-859.
- Yang, Nan, MacArthur, Daniel G, Gulbin, Jason P, Hahn, Allan G, Beggs, Alan H, Easteal, Simon, & North, Kathryn. (2003). ACTN3 genotype is associated with human elite athletic performance. *The American Journal of Human Genetics*, 73(3), 627-631.